

BEST AVAILABLE COPY

source 1..2556
/organism="unknown"
BASE COUNT 797 a 614 c 514 g 631 t
ORIGIN

Query Match 99.1% Score 2534; DB 6; Length 2556;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GGGAGCCCTGTGACCAAAAGTGGCTCCGGTCAAGCCGCCCTCTTAAAGTCAGAGGGCCCT 60
DB 1 GGGAGCCCTGTGACCAAAAGTGGCTCCGGTCAAGCCGCCCTCTTAAAGTCAGAGGGCCCT 60
QY 61 AGGCTGCTGCTCTCTAGATAGTGGCCGTGAAGCCCAACACACAGACATCACTT 120
DB 61 AGGCTGCTGCTCTCTAGATAGTGGCCGTGAAGCCCAACACACAGACATCACTT 120
QY 121 CCTGCTAATTTGACGCTTCTCCAGAAACGTTTGAATTAAGTACAGTGGCCCTTG 180
DB 121 CCTGCTAATTTGACGCTTCTCCAGAAACGTTTGAATTAAGTACAGTGGCCCTTG 180
QY 181 ATGTGTGATCTCCTCAGCAAACTGTAAACAAGCCGAGACAGATTAACATTAACCTCA 240
DB 181 ATGTGTGATCTCCTCAGCAAACTGTAAACAAGCCGAGACAGATTAACATTAACCTCA 240
QY 241 AGCCAGCAGTACAGCCGAATCTCTCAAAAGTCAAAATCTGTACAGTGGCCGAACTTAC 300
DB 241 AGCCAGCAGTACAGCCGAATCTCTCAAAAGTCAAAATCTGTACAGTGGCCGAACTTAC 300
QY 301 TCACATTAATTCAGAAAGTGGCAGTACACCTTTAAAAATGGCAAAATAGAACT 360
DB 301 TCACATTAATTCAGAAAGTGGCAGTACACCTTTAAAAATGGCAAAATAGAACT 360
QY 361 ACTGTGTAAACACTGTCTCCGAAGCCCTCTCACTACAACTGTGGCTGTGCCAACCACT 420
DB 361 ACTGTGTAAACACTGTCTCCGAAGCCCTCTCACTACAACTGTGGCTGTGCCAACCACT 420
QY 421 RGTGTCACAGTTTACCTCCTGGAAAGCCATGTGAATCTGTAACTACCTGTGAAGCTTAA 480
DB 421 -GTGTGACAGTTTACCTCCTGGAAAGCCATGTGAATCTGTAACTACCTGTGAAGCTTAA 479
QY 481 TTTGGGAGCATCACTCCTCTTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 540
DB 481 TTTGGGAGCATCACTCCTCTTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 539
QY 541 TGTTCAGATTAATCTTCTCCGACAACTGTAGAAATGTGAAGAAATGCAAACTTCT 600
DB 541 TGTTCAGATTAATCTTCTCCGACAACTGTAGAAATGTGAAGAAATGCAAACTTCT 599
QY 601 TGGAAATTTAATAAATATACATGTAGTATGATCACTCCCTGTGAATGGGGCAAAATGT 660
DB 601 TGGAAATTTAATAAATATACATGTAGTATGATCACTCCCTGTGAATGGGGCAAAATGT 659
QY 661 GAAGAAGCTGTGGAACAACCTTTGATGCAAAAATGGAAGCAGAAATTTATAGAA 720
DB 661 GAAGAAGCTGTGGAACAACCTTTGATGCAAAAATGGAAGCAGAAATTTATAGAA 719
QY 721 ACTGTATGTTGAACCTCAAGTCTTCACTCAAGCCCTCACTGCTGTTCTTTTGAAGAAAG 780
DB 721 ACTGTATGTTGAACCTCAAGTCTTCACTCAAGCCCTCACTGCTGTTCTTTTGAAGAAAG 779
QY 781 CGTGTGTCCTTACAGAACTTCTGCTTACCTCCAGAGCTTCACTCCAGAAATGTGTCA 840
DB 781 CGTGTGTCCTTACAGAACTTCTGCTTACCTCCAGAGCTTCACTCCAGAAATGTGTCA 839
QY 841 GCAGACTTCTAGTACATGCTCATTTGTACTGTACTACAAAGTAACTCACTTCTCTGT 900
DB 841 GCAGACTTCTAGTACATGCTCATTTGTACTGTACTACAAAGTAACTCACTTCTCTGT 899
QY 901 GGTGACAACTACAGTGTCTCAAGCCAGTCTGAAGAAATCAATTAATTTCTGTGAGCAAC 960
DB 901 GGTGACAACTACAGTGTCTCAAGCCAGTCTGAAGAAATCAATTAATTTCTGTGAGCAAC 959

QY 961 AGCACCAGAACTGTGTCAAGTCAAACTTTGAACCCACTTGTGTCAAGTGGAGCAAA 1020
DB 961 AGCACCAGAACTGTGTCAAGTCAAACTTTGAACCCACTTGTGTCAAGTGGAGCAAA 1019
QY 1021 AGCTGAGTTGTGACACTTCACTTCTGTGGGCCCAACTGTGTCAAGAGGAAACACAGC 1080
DB 1021 AGCTGAGTTGTGACACTTCACTTCTGTGGGCCCAACTGTGTCAAGAGGAAACACAGC 1079
QY 1081 TGGAACTGTTGCTTCACTTCAAAACCACTTGTGATCATGTGTGCAAAACAGATGAC 1140
DB 1081 TGGAACTGTTGCTTCACTTCAAAACCACTTGTGATCATGTGTGCAAAACAGATGAC 1139
QY 1141 CACGCTCTACATGCACTGAAAGCACTTGTGTGATCATGTGTGCAAAACAGATGAC 1200
DB 1141 CACGCTCTACATGCACTGAAAGCACTTGTGTGATCATGTGTGCAAAACAGATGAC 1199
QY 1201 TCCAGAGTAACTTTGAGAAACCTTCAAGTGTGAGTATTTCTTCCATCTGTGAACC 1260
DB 1201 TCCAGAGTAACTTTGAGAAACCTTCAAGTGTGAGTATTTCTTCCATCTGTGAACC 1259
QY 1261 TGTGTCTCTCTGCTGTGGACACATCTGCAAGCCCTGTATTTGGGACTCCAGTTCAAT 1320
DB 1261 TGTGTCTCTCTGCTGTGGACACATCTGCAAGCCCTGTATTTGGGACTCCAGTTCAAT 1319
QY 1321 CAAACTTGGCCAGCCGGGCCCTGTCTTCAACAACAGCTGGGATTCCAACAGGCACTTC 1380
DB 1321 CAAACTTGGCCAGCCGGGCCCTGTCTTCAACAACAGCTGGGATTCCAACAGGCACTTC 1379
QY 1381 AACCAAGCAACTTCTCTGATTTGTTACGTAGTGTGACAGCCCTTCAAGAGGCAATTAATA 1440
DB 1381 AACCAAGCAACTTCTCTGATTTGTTACGTAGTGTGACAGCCCTTCAAGAGGCAATTAATA 1439
QY 1441 ACAAGTGAACCAATTTCACTTCTGCAACATTTGACATTCAGAAATGGGACAGAGAAC 1500
DB 1441 ACAAGTGAACCAATTTCACTTCTGCAACATTTGACATTCAGAAATGGGACAGAGAAC 1499
QY 1501 GATGCCAGTGAACACATTAATACCTAAGTACAGTTCCTTCAAGCTTCTCAATTTAAAGA 1560
DB 1501 GATGCCAGTGAACACATTAATACCTAAGTACAGTTCCTTCAAGCTTCTCAATTTAAAGA 1559
QY 1561 AATTTACCTGCTGTGAATTAATCTGTCACTTCAACATCTCTTACTGAGAAATAA 1620
DB 1561 AATTTACCTGCTGTGAATTAATCTGTCACTTCAACATCTCTTACTGAGAAATAA 1618
QY 1621 GAATTAAGAGAAATGTATACATGCTTCCGAGATGAGATGACATCAATGATGTACTT 1680
DB 1621 GAATTAAGAGAAATGTATACATGCTTCCGAGATGAGATGACATCAATGATGTACTT 1678
QY 1681 CTATGCGAGGGGTCAACCTTAATGAAGAAATGCTGCACTTTAGACAAACTCTGAAT 1740
DB 1681 CTATGCGAGGGGTCAACCTTAATGAAGAAATGCTGCACTTTAGACAAACTCTGAAT 1738
QY 1741 TGGTGGCAGACATCACTTCACTGATGAAGTGAAGAACATTTCTTTTATTTGAGACTTAC 1800
DB 1741 TGGTGGCAGACATCACTTCACTGATGAAGTGAAGAACATTTCTTTTATTTGAGACTTAC 1798
QY 1801 AAAAGAGAACTTTAGACATTTGTTAAAGACATGACATTAACAGAACTTCACTGTGATG 1860
DB 1801 AAAAGAGAACTTTAGACATTTGTTAAAGACATGACATTAACAGAACTTCACTGTGATG 1858
QY 1861 TGAACCTTGAATCTCCAGACAGCAAGAGCACTACAGAGCCCTTCTAGAAAAACTGACTG 1920
DB 1861 TGAACCTTGAATCTCCAGACAGCAAGAGCACTACAGAGCCCTTCTAGAAAAACTGACTG 1918
QY 1921 CAATGCTCAGAGTGAAGTACTTCAAGGCAAGTAAATTAATCACTCTGTGTAGAG 1980
DB 1921 CAATGCTCAGAGTGAAGTACTTCAAGGCAAGTAAATTAATCACTCTGTGTAGAG 1978
QY 1981 ATACAGGTCAAGCTCAATTTCTTGAAGAACTGGATCAATTTGGAGAGCAGAGAAAG 2040
DB 1981 ATACAGGTCAAGCTCAATTTCTTGAAGAACTGGATCAATTTGGAGAGCAGAGAAAG 2038
QY 2041 AATTGGAAGAAAGAAATGTTACTTAAGGACCAAGAGCTGTTCTTAATAAGAAAGATC 2100

RESULT 2		HSTAFII	
LOCUS	HSTAFII	2556 bp	mRNA
DEFINITION	H.sapiens TAFII105 mRNA, partial.	linear	PRI 14-NOV-1996
ACCESSION	Y09321		
VERSION	Y09321.1	GI:1659688	
KEYWORDS	TAFII105 gene.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota, Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2556) Dikstein, R., Zhou, S., and Tjian, R.		
TITLE	Human TAFII 105 is a cell type-specific TFIID subunit related to hTFII130		
JOURNAL	Cell 87 (1), 137-146 (1996)		
MEDLINE	97011146		
PUBMED	8658156		
REFERENCE	2 (bases 1 to 2556)		
AUTHORS	Dikstein, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-NOV-1996) R. Dikstein, Weizmann Institute of Science, Dept. Biochemistry, Rehovot, 76100, ISRAEL		
FEATURES	Location/Qualifiers		
SOURCE	1..2556		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="Hela"		
	1..2406		
	/gene="TAFII105"		
CDS	<1..2406		

Query Match	Best Local Similarity	Score	DB	Length
Matches	Conservative	Mismatches	Indels	Gaps
QY 1	GGGACCCGTGGTACCAAGTGGCTCCGGTCAGCGCCCTCTTAAAGTACGACGGCCCT	99.1%	2534	2556
Db 1	GGGACCCGTGGTACCAAGTGGCTCCGGTCAGCGCCCTCTTAAAGTACGACGGCCCT	99.9%	0	0
QY 61	AGGCGCCTGCTCCTCAGATAGTGGCCGTGAAGCCCCCAACACGACACATTCAGTTT	0	0	2
Db 61	AGGCGCCTGCTCCTCAGATAGTGGCCGTGAAGCCCCCAACACGACACATTCAGTTT	0	0	2
QY 121	CCTGCTAATTTGGACACTCTCCAGAAACCCCTTTTGGATTAAGAAGAAAGAGGTCCTTG	0	0	0
Db 121	CCTGCTAATTTGGACACTCTCCAGAAACCCCTTTTGGATTAAGAAGAAAGAGGTCCTTG	0	0	0
QY 181	ATGTGGTATCTCCTCAGCAAACTGTATAAGAGCCGAGACACAAAGTACATACCTCA	0	0	0
Db 181	ATGTGGTATCTCCTCAGCAAACTGTATAAGAGCCGAGACACAAAGTACATACCTCA	0	0	0
QY 241	AGGCGACAGTACGAGGAAATCTCAAAACATCAAAATCTTACAGTCCGACACTTAC	0	0	0
Db 241	AGGCGACAGTACGAGGAAATCTCAAAACATCAAAATCTTACAGTCCGACACTTAC	0	0	0
QY 301	TCACAATTTAATCAAGAAAGTGGCAGTGCACCTGTGTAAAAAATTTGGCACAATAGGA	0	0	0
Db 301	TCACAATTTAATCAAGAAAGTGGCAGTGCACCTGTGTAAAAAATTTGGCACAATAGGA	0	0	0
QY 361	ACTGTGGTAACCACTGTTCCGAAACCTTCTCAGTACATCTGTGGCTGGCCAAACG	0	0	0
Db 361	ACTGTGGTAACCACTGTTCCGAAACCTTCTCAGTACATCTGTGGCTGGCCAAACG	0	0	0
QY 421	RGTTGTACAGTACTCTCTGGAAAGCCATTGTAATAGTAACTGACCTGGAAGCCCTTCA	0	0	0
Db 421	RGTTGTACAGTACTCTCTGGAAAGCCATTGTAATAGTAACTGACCTGGAAGCCCTTCA	0	0	0
QY 481	TTTGGAGACATCAATCCATCTTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAG	0	0	0
Db 481	TTTGGAGACATCAATCCATCTTCAAAATGAGCCCAATCTTAAAGCAGAGAACTCAG	0	0	0
QY 541	TGTTCAAGTTAATCTTCTCCGAAAGTGAAGCAAAATGTGAAGAAATGCAAGAACTTCT	0	0	0
Db 541	TGTTCAAGTTAATCTTCTCCGAAAGTGAAGCAAAATGTGAAGAAATGCAAGAACTTCT	0	0	0
QY 601	TGCAATGTATAAATAGCATGTAGTGAATCAGTCCCTGAAATGGGCGAAAAATGT	0	0	0
Db 601	TGCAATGTATAAATAGCATGTAGTGAATCAGTCCCTGAAATGGGCGAAAAATGT	0	0	0
QY 661	GAAAGACCTGGTGAACAACCTTTGGATGCAAAAATGGAAGCAGAAAGAAATTTACTAG	0	0	0
Db 661	GAAAGACCTGGTGAACAACCTTTGGATGCAAAAATGGAAGCAGAAAGAAATTTACTAG	0	0	0
QY 721	ACTATATGTGAAGTCAAGTCTCAGCTCAGCTCAGCTCTTCTTCTTAAGAAAAG	0	0	0
Db 721	ACTATATGTGAAGTCAAGTCTCAGCTCAGCTCAGCTCTTCTTCTTAAGAAAAG	0	0	0

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 21.8929 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GILVTVKAVASAPKRVSSGP.....KZHOHERALFTIRILLITLY 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*\n2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*\n3: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*\n4: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*\n5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*\n6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3990	93.6	801	1	US-08-725-012-2
2	1307.5	30.7	737	1	US-08-188-582-16
3	1307.5	30.7	737	1	US-08-646-715-16
4	775	18.2	921	1	US-08-188-582-2
5	775	18.2	921	1	US-08-646-715-2
6	237	5.6	2035	1	US-08-046-585-5
7	237	5.6	2035	1	US-08-393-703-5
8	237	5.6	2035	5	PCT-US93-11721-5
9	198	4.6	862	1	US-08-325-267A-4
10	197	4.6	2972	4	US-09-579-181-2
11	197	4.6	3118	4	US-09-579-181-1
12	191.5	4.5	894	3	US-08-362-525-22
13	191.5	4.5	894	3	US-08-971-692-15
14	184.5	4.3	1537	1	US-08-325-267A-2
15	178	4.2	1721	3	US-08-700-651-5
16	178	4.2	1721	3	US-08-928-361B-6
17	175.5	4.1	3569	4	US-08-061-376-5
18	172	4.0	750	4	US-09-165-239A-4
19	168	3.9	1837	3	US-08-928-361B-5
20	166	3.9	805	4	US-09-103-429A-4
21	164.5	3.9	752	1	US-08-244-189-2
22	164	3.8	903	2	US-08-853-310-2
23	160	3.8	1125	4	US-09-513-783A-152
24	160	3.8	1610	4	US-09-513-783A-22
25	158.5	3.7	786	4	US-09-103-429A-3
26	157.5	3.7	941	4	US-07-757-022B-14
27	157.5	3.7	1022	4	US-07-757-022B-84

28	157.5	3.7	1038	4	US-07-757-022B-74	Sequence 74, Appl
29	157.5	3.7	1049	4	US-07-757-022B-58	Sequence 58, Appl
30	157.5	3.7	1140	4	US-07-757-022B-104	Sequence 104, Appl
31	157.5	3.7	1270	4	US-07-757-022B-44	Sequence 44, Appl
32	157.5	3.7	1311	4	US-07-757-022B-42	Sequence 42, Appl
33	157.5	3.7	1313	4	US-07-757-022B-142	Sequence 142, Appl
34	157.5	3.7	1314	4	US-07-757-022B-46	Sequence 46, Appl
35	157.5	3.7	1320	4	US-07-757-022B-50	Sequence 50, Appl
36	157.5	3.7	1320	4	US-07-757-022B-60	Sequence 60, Appl
37	157.5	3.7	1354	4	US-07-757-022B-48	Sequence 48, Appl
38	157.5	3.7	1361	4	US-07-757-022B-40	Sequence 40, Appl
39	157.5	3.7	1363	4	US-07-757-022B-52	Sequence 52, Appl
40	157.5	3.7	1404	4	US-07-757-022B-2	Sequence 2, Appl
41	157.5	3.7	1404	4	US-07-757-022B-62	Sequence 62, Appl
42	157	3.7	2843	1	US-07-741-940-2	Sequence 2, Appl
43	157	3.7	2843	1	US-08-289-548A-2	Sequence 2, Appl
44	157	3.7	2843	1	US-08-452-65A-2	Sequence 2, Appl
45	157	3.7	2843	2	US-08-370-235A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-725-012-2
Sequence 2, Application US/08725012
Patent No. 5710025
GENERAL INFORMATION:
APPLICANT: Dikstein, Rivka
APPLICANT: Tjian, Robert
TITLE OF INVENTION: B-Cell Specific Transcription Factor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-012-2
Query Match 93.6% Score 3990; DB 1; Length 801;
Best Local Similarity 99.9% Pred. No. 9.6e-319;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GILVTVKAVASAPKRVSSGPLPAPQIVAVKAPNTTIOFPANTLOPPTVILIKNSGFL 60
Db 1 GILVTVKAVASAPKRVSSGPLPAPQIVAVKAPNTTIOFPANTLOPPTVILIKNSGFL 60
QY 61 MAVSPQQTVAETRAETSNITSRPAVPANPQTVKICVVPNSSQLIKKVAATPVKKLAQIGT 120

Db 61 MLVSPQOQTVTAETTSNITSRPAPVAPNPQVKTCTVPPNSSQILKKAATVPYKLAQIGT 120
 QY 121 TVTTPVPPSSVQSAVAPPTSVTVTPGKPLNTVTTLTKSSLSGASSTPSNEPRLKAENSA 180
 Db 121 TVTTPVPPSSVQSAVAPPTSVTVTPGKPLNTVTTLTKSSLSGASSTPSNEPRLKAENSA 180
 QY 181 VOINLSPMLBNVKKCNFLMLIKLACSGSQSPMGONVKKVQLLDATIEAEFPK 240
 Db 181 VOINLSPMLBNVKKCNFLMLIKLACSGSQSPMGONVKKVQLLDATIEAEFPK 240
 QY 241 LYLVLKSSPOPHLVPFLKKSVALRQLLPNSQSFTQQCVQOTSSDMVATCTTTVTTS 300
 Db 241 LYLVLKSSPOPHLVPFLKKSVALRQLLPNSQSFTQQCVQOTSSDMVATCTTTVTTS 300
 QY 301 VTTTSSSQSEKSIIVSGATPRVSVQTNPLAGPVGAKGVVTLHSVGPAAATGTTA 360
 Db 301 VTTTSSSQSEKSIIVSGATPRVSVQTNPLAGPVGAKGVVTLHSVGPAAATGTTA 360
 QY 361 GTGLQTSKPLVTSVANTVTTVSLQPKPVVSGTAVTSLPATEFGESGAATCLPSVKP 420
 Db 361 GTGLQTSKPLVTSVANTVTTVSLQPKPVVSGTAVTSLPATEFGESGAATCLPSVKP 420
 QY 421 VVSPCMHICKPVYGTVOYKLAQPGVLSOPAGIPGSSSKQLSLFHVVOQPSGNEK 480
 Db 421 VVSPCMHICKPVYGTVOYKLAQPGVLSOPAGIPGSSSKQLSLFHVVOQPSGNEK 480
 QY 481 QVTTTSSSTLTIOCKGCKMPVNTIIPTSOPASTLKQITLPKNTILSLQASPTOKN 540
 Db 481 QVTTTSSSTLTIOCKGCKMPVNTIIPTSOPASTLKQITLPKNTILSLQASPTOKN 540
 QY 541 IKENVTSCFDEDEDINDVTSMAGVNLNEENACILATNSEVLTILQSCKEPFLFGALQ 600
 Db 541 IKENVTSCFDEDEDINDVTSMAGVNLNEENACILATNSEVLTILQSCKEPFLFGALQ 600
 QY 601 KRLLDGGKHDITELNSDANVLSOATQERLGLKELVIAQHRMTTYKASNYILCS 660
 Db 601 KRLLDGGKHDITELNSDANVLSOATQERLGLKELVIAQHRMTTYKASNYILCS 660
 QY 661 TRSOLKFLKLDLEKORKEERMLKAKSRNKEDPOLRLQKAKELQQLAQT 720
 Db 661 TRSOLKFLKLDLEKORKEERMLKAKSRNKEDPOLRLQKAKELQQLAQT 720
 QY 721 QHRDANLITTAIGPKKRPLESIGELKDNILASGTSSLTATQOLRRPRTICRLDI 780
 Db 721 QHRDANLITTAIGPKKRPLESIGELKDNILASGTSSLTATQOLRRPRTICRLDI 780
 QY 781 FCMEOERKYSRALTALIK 801
 Db 781 FCMEOERKYSRALTALIK 801

RESULT 2
 US-08-188-582-16
 Sequence 16, Application US/08188582
 Patent No. 5534410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
 APPLICANT: Comai, Lucio
 APPLICANT: Dynlacht, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoko
 APPLICANT: Wang, Edith
 APPLICANT: Weinzierl, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 NUMBER OF INVENTIONS: NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE
 CORRESPONDENCE ADDRESS: 36
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA

ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/188,582
 FILING DATE: 28-JAN-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A.
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 737, amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-188-582-16

Query Match 30.7%; Score 1307.5; DB 1; Length 737;
 Best Local Similarity 40.0%; Pred. No. 1,3e-98;
 Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;

QY 13 PKRVSSG-----RLPAPQIVAAKAPNTTIOBPANLQIPGVLLKNSGRLM 61
 Db 68 PTAATSGIRATLPPVYALRPLP-----PONPNIQ--NFQLPKGVIVSENGQL 118
 QY 62 LVSPOQTVTR-----AETSNITSRPAPVAPNPQVKTCTVPPNSSQILKKAATVPYK 115
 Db 119 MT-POOALQMOQAHAQOQTMAPRPAPTSAAPVQISTVQAPGPIIAR-QVTP----- 172
 QY 116 AQIGTTVTVTPKPSVQSAVAPPTSVTVTPGKPLNTYT--TLKSSLSGASST--TPS 168
 Db 173 ----TTIIKQV---SOAQTVQPSATLQPSQVOPOLVGGAAQTAAGTAAVQGTCTP 225
 QY 169 NE-PULKAENSAVOINLSPMLBNVKKCNFLMLIKLACSGSQSPMGONVKKVQL 227
 Db 226 RTVPGATTTSSAAT-----TMENVKCKKNFLSTLILKLAASSGKOSTTAANVKELVONL 279
 QY 228 LDKATEAEFPRLKLVLEKSSPOPHLVPFLKKSVALRQLLPNSQSFTQQCVQO-----TS 283
 Db 280 LDGKTEADEFTSRLYRELNSSPQPLVPFLKRSPLALROLTPDASAFTQQSQOQPPPTS 339
 QY 284 SDMVATCTTTVTTSPPVTTTSSSQSEKSIIVSGATPRVSVQTNPLAGPVGAKGV 343
 Db 340 Q-----ATTALRAVAVLSSSVQRTAGKTAATVTSALQPVLSL----- 376
 QY 344 VTLHSVGPAAATGTTAGTGLLQTSKPLVTSVANTVTTVSLQPKPVVSGTAVTSLP 403
 Db 377 -----TQPTQGVGKQGPPLVLIQ-----QPKRP-----GALIRPQV 410
 QY 404 TFEENSGAATCLPSVKPVVSCMCHICKPVYGTVOYKLAQPGVLSOPAGIPGSSSKQ 463
 Db 411 TLTOT-----PVALRQPH-NRIMLTTPQIOL----- 437
 QY 464 LSLFHVVOQPSGNEKQVTTTSSSTLTIOCKGCKMPVNTIIPTSOPASTLKQITL 523
 Db 438 ----- 451
 QY 524 PGNKIL---SLQASPTOKNRIKENVTSCFDEDEDINDVTSMAGVNLNEENACILATSEL 580
 Db 452 PGFALSAVSAQAAAKNKLKEKGGSPRDDINDVASMAGVNLSEESARILATSEL 511
 QY 581 VGTLLQSCKEPFLFGALQRLIDGKHDITELNSDANVLSOATQERLGLKELKTA 640
 Db 512 VGTLLQSCKEPFLFGALQRLIDGKHDITELNSDANVLSOATQERLGLKELKTA 640

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 16.4197 Seconds
(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GTLVTKVAPVSAAPKVVSGP.....KZHOERALTITLTLTLTY 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3990	93.6	801	T2DT_HUMAN	Q92750 homo sapien
2	1307.5	30.7	1083	T2D3_HUMAN	000268 homo sapien
3	775	18.2	921	T2D3_DROME	P47825 drosophila
4	239.5	5.6	2035	HFCL_HUMAN	P51810 homo sapien
5	238	5.6	2090	AMEY_HUMAN	P51611 mesocricetu
6	220	5.2	1367	AMEY_YEAST	P08640 saccharomyc
7	219	5.1	5179	MUC2_HUMAN	002817 homo sapien
8	213.5	5.0	5376	ZAN_MOUSE	008799 mus musculu
9	206.5	4.8	865	CPN_DROME	002910 drosophila
10	205.5	4.8	1161	DAN4_YEAST	P47179 saccharomyc
11	203.5	4.8	1322	YAG3_YEAST	P39712 saccharomyc
12	201	4.7	725	AGAL_YEAST	P33232 saccharomyc
13	199	4.7	2700	ZAN_HUMAN	Q9Y493 homo sapien
14	197.5	4.6	881	YUHE_YEAST	P47033 saccharomyc
15	197.5	4.6	1609	FTG_YEAST	P25653 saccharomyc
16	193.5	4.5	670	VG50_HSV1	000130 icetaliuid h
17	191.5	4.5	1260	ALST_CANAL	P46300 candida alb
18	191	4.5	797	VGX_HSV1	P28968 equine herp
19	187.5	4.4	3726	TRX_DROME	P20659 drosophila
20	186	4.4	3178	TS89_CAEEL	Q09624 caenorhabdi
21	184.5	4.3	1537	FLO1_YEAST	P32768 saccharomyc
22	182	4.3	1075	FLOS_YEAST	P38894 saccharomyc
23	182	4.3	3866	HRX_MOUSE	P35200 mus musculu
24	181.5	4.3	662	MUC1_XENLA	Q05049 xenopus lae
25	181.5	4.3	1199	N121_RAT	P52591 rattus norv
26	181	4.2	1858	YK96_DICDI	004893 saccharomyc
27	177.5	4.2	1140	YK96_YEAST	Q9Y490 homo sapien
28	176	4.1	2541	TAL1_HUMAN	P33658 homo sapien
29	175.5	4.1	2090	N214_HUMAN	Q03164 homo sapien
30	175.5	4.1	3969	HRX_HUMAN	P41809 saccharomyc
31	173.5	4.1	1802	HRK1_YEAST	Q95X47 mus musculu
32	173	4.1	2404	SON_MOUSE	Q07284 epstein-bar
33	172	4.0	886	VGP3_BBYA8	

34	172	4.0	1018	HMW1_MYCPN	O50365 mycoplasma
35	172	4.0	1119	ALS3_CANAL	Q74623 candida alb
36	172	4.0	1169	YK82_YEAST	P36170 saccharomyc
37	171.5	4.0	1780	YK26_CAEEL	P34333 caenorhabdi
38	170	4.0	1306	MSB2_YEAST	P34334 saccharomyc
39	169.5	4.0	388	MPY1_YEAST	P50105 saccharomyc
40	169.5	4.0	745	OCT1_PIG	Q29076 sus scrofa
41	169.5	4.0	2426	SON_HUMAN	P16583 homo sapien
42	169	4.0	1229	N121_HUMAN	Q9Y293 homo sapien
43	166.5	3.9	743	OCT1_HUMAN	P14659 homo sapien
44	165	3.9	606	SP2_HUMAN	Q02086 homo sapien
45	165	3.9	2541	TAL1_MOUSE	P26039 mus musculu

ALIGNMENTS

```

RESULT 1
ID T2DT_HUMAN STANDARD: PRT; 801 AA.
AC 092750:
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor TFIIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_Taxid:9606;
RN [1]
RP MEDLINE=97011146; PubMed=8858156;
RX Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIIID subunit related to
RT hTAFII130.";
RT Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIIID THAT MAY FUNCTION AS
CC A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIIID IS A
CC MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
CC PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIIID IS COMPOSED OF DATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
CC HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Y09321; CAAT0499.1; -
CC Genew; HGNC:11538; TAF4B.
CC MIM: 601689; -
CC Interpro: IPR003894; TAF_hom.
CC SMART: SM00549; TAFH; 1.
CC KJ Transcription regulation; Nuclear protein.
CC NON_TER 1
CC FT SEQUENCE 801 AA; 85658 MW; D12B4932FEA9CD2 CRC64;
CC SQ
Query Match 93.6%; Score 3990; DB 1; Length 801;
Best Local Similarity 99.9%; Pred. No. 5.1e-199;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 GTLVTKVAPVSAAPKVVSGPRLAPQIVYAKAPVTTTIOPRANLOPGTVLIKSNSGPL 60
|||||

```

Db 1 GTTATKAPASAPKPVSSGRLPAPQIVAKAPNTTIOFANLQLPCTVLKNSGRL 60
 QY 61 MLYSPQOYTRAEETTSNITSRAVPAVPAVQVAVICVYVNSSQLIKKVAVTPEVKAQIGT 120
 Db 61 MLYSPQOYTRAEETTSNITSRAVPAVPAVQVAVICVYVNSSQLIKKVAVTPEVKAQIGT 120
 QY 121 TTTTTPKPVSSQVAVPVSVTATVPGKRLNVTTLKPSLGLASSPSEMPKLNKNSA 180
 Db 121 TTTTTPKPVSSQVAVPVSVTATVPGKRLNVTTLKPSLGLASSPSEMPKLNKNSA 180
 QY 181 VOINLSPTMLEVKKCKNKLAMLIKACSGSSPEMGQVKKVLYVQLDLAKTEAEFFTRK 240
 Db 181 VOINLSPTMLEVKKCKNKLAMLIKACSGSSPEMGQVKKVLYVQLDLAKTEAEFFTRK 240
 QY 241 LYVELKSSPQPLVPLFKSVVALRQLLPNSQSFIOOCVQOVSQDMVATCTTCTTTSVP 300
 Db 241 LYVELKSSPQPLVPLFKSVVALRQLLPNSQSFIOOCVQOVSQDMVATCTTCTTTSVP 300
 QY 301 VTTTSSSSQSEKSIIVSGATAPRTVSVQTLNLAQVYKAAVYVTLHSGPFAAGGTTA 360
 Db 301 VTTTSSSSQSEKSIIVSGATAPRTVSVQTLNLAQVYKAAVYVTLHSGPFAAGGTTA 360
 QY 361 GTGLQTSKPLTVSVANVTYVLSLOPEKPVYSGTAVTSLPVTGEGTGAICLPVYK 420
 Db 361 GTGLQTSKPLTVSVANVTYVLSLOPEKPVYSGTAVTSLPVTGEGTGAICLPVYK 420
 QY 421 VVSPFMDHCKVIGTVPQVQIKLAQPGPVLSQAGIPPTGSSSKQLPFLHVYVQPSGNEK 480
 Db 421 VVSPFMDHCKVIGTVPQVQIKLAQPGPVLSQAGIPPTGSSSKQLPFLHVYVQPSGNEK 480
 QY 481 QVTTTSHSTLTIOCGCGTQMPVNTIIPTSOPPASILKQTLIPNKLTLSQASPTOKNR 540
 Db 481 QVTTTSHSTLTIOCGCGTQMPVNTIIPTSOPPASILKQTLIPNKLTLSQASPTOKNR 540
 QY 541 IKENVTSCFRDNDINDVYSNAGVNLNENACILATNSLVTGLIQSCDEPFLFGLAQ 600
 Db 541 IKENVTSCFRDNDINDVYSNAGVNLNENACILATNSLVTGLIQSCDEPFLFGLAQ 600
 QY 601 KRIDIGKHHITELNSDAVNLISQATQRLGLLEKLTALQHRMTYTKASENTIILSD 660
 Db 601 KRIDIGKHHITELNSDAVNLISQATQRLGLLEKLTALQHRMTYTKASENTIILSD 660
 QY 661 TRSOLKFLKDLQLEKORDEERELKKAASRSNKEDPEQLRKOKAKELQOELAQI 720
 Db 661 TRSOLKFLKDLQLEKORDEERELKKAASRSNKEDPEQLRKOKAKELQOELAQI 720
 QY 721 QHRDANLTATAAIGPRKKRPLESGIEGLKDNLASGTSSTATKQLHRPRITRICALDLI 780
 Db 721 QHRDANLTATAAIGPRKKRPLESGIEGLKDNLASGTSSTATKQLHRPRITRICALDLI 780
 QY 781 FCMEOEREMKYSRALYIALK 801
 Db 781 FCMEOEREMKYSRALYIALK 801

RESULT 2
 T2D3_HUMAN
 ID T2D3_HUMAN STANDARD: PRT: 1083 AA.
 AC 000268; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)
 GN (TAFII135) (TAFII-130) (TAFII130).
 OS TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OC Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97336072; PubMed=9192867;
 RA Mengus G., May M., Carre L., Chambon P., Davidson I.;

RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
 RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
 RT mammalian cells.";
 RL Genes Dev. 11:1381-1395(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavridis G., Almeida J.P., Babbage A.R., Baguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gylliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekoshts K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Lehtvala M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McKay K., McMurtry A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RA "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=97098442; PubMed=8942982;
 RA Taese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
 RA "Molecular cloning and analysis of two subunits of the human TFIID
 RT complex: hTAFII130 and hTAFII100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
 CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Y11354; CAI2189.1; -
 CC EMBL: AL137077; CAC36006.1; -
 CC EMBL: AL109911; CAC22312.2; -
 CC EMBL: U75308; AAC50901.1; -
 CC TRANSFAC: T02328; -
 CC GeneW: HGNC:11537; TAF4.
 DR MIM: 601796; -
 DR InterPro: IPR003894; TAF_hom.
 DR SMART: SM00549; TAFH; 1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 39 42 POLY-HIS.
 FT DOMAIN 52 57 POLY-ALA.
 FT DOMAIN 98 101 POLY-GLY.
 FT DOMAIN 142 148 POLY-ALA.
 FT DOMAIN 268 275 POLY-PRO.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:29:40 : Search time 6475 Seconds

(without alignments)
11497.294 Million cell updates/sec

Title: US-09-763-909-1

Sequence: 1 999accctgtgacccaagt.....cttataactcttacctat 2558

Scoring table: IDENTITY_NNC
Gapop 10.0, Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

GenBank: 1: gb_ba: 2: gb_ba: 3: gb_ba: 4: gb_ba: 5: gb_ba: 6: gb_ba: 7: gb_ba: 8: gb_ba: 9: gb_ba: 10: gb_ba: 11: gb_ba: 12: gb_ba: 13: gb_ba: 14: gb_ba: 15: gb_ba: 16: gb_ba: 17: gb_ba: 18: gb_ba: 19: gb_ba: 20: gb_ba: 21: gb_ba: 22: gb_ba: 23: gb_ba: 24: gb_ba: 25: gb_ba: 26: gb_ba: 27: gb_ba: 28: gb_ba: 29: gb_ba: 30: gb_ba: 31: gb_ba: 32: gb_ba: 33: gb_ba: 34: gb_ba: 35: gb_ba: 36: gb_ba: 37: gb_ba: 38: gb_ba: 39: gb_ba: 40: gb_ba: 41: gb_ba:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2534	99.1	2556	6	181227
2	2534	99.1	2556	6	HSTAF11
3	1145.2	44.8	182805	2	AC017007
4	580	22.7	119081	2	AC121320
5	580	22.7	169240	9	AC022069
6	580	22.7	190782	2	AP002752
7	338.2	13.2	3603	6	123466
8	338.2	13.2	3603	6	145759
9	337.2	13.2	3252	9	HSU75308
10	337.2	13.2	3252	9	HSTAF113
11	325.8	12.7	2156	10	AY038601
12	294.6	11.5	174222	2	AP001197
13	294.6	11.5	175553	2	AP001096
14	294.6	11.5	182884	9	AC007996
15	294.6	11.5	190782	2	AP002752
16	148.6	5.8	333300	2	AC125091
17	148.6	5.8	16903	2	AC016839
18	148.6	5.8	61682	2	AC118057
19	147	5.7	277892	2	AC125105
20	146.6	5.7	205466	3	AC127767
21	143.8	5.6	33535	3	563550
22	142.2	5.6	3745	3	AY069807
23	142.2	5.6	4615	3	DR0FAF110X
24	142.2	5.6	4615	6	123459
25	142.2	5.6	4615	6	145752
26	137.6	5.4	205466	2	AC127767
27	124.4	4.9	62974	2	AC118056
28	121.8	4.8	277892	2	AC125105
29	112.4	4.4	158509	2	AC114177
30	109.8	4.3	62974	2	AC118056
31	99	3.9	158509	2	AC114177
32	98.4	3.8	191918	2	AC102692
33	94.6	3.7	121982	2	HS1107624
34	92	3.6	333300	2	AC125091
35	91.2	3.6	64035	2	AC119227
36	89.6	3.5	194426	2	AL663067
37	71	2.8	112277	2	AC113239
38	62.4	2.4	2147	9	AK097744
39	58.6	2.3	59665	2	AC014422
40	58.6	2.3	168479	3	AC093454
41	58.6	2.3	176056	3	AC010066
42	58.6	2.3	283821	3	AE003528
43	57.6	2.3	7218	6	166494
44	56.8	2.2	126712	2	AL137077
45	56.4	2.2	201470	2	AC113649

ALIGNMENTS

RESULT 1
LOCUS 181227 2556 bp DNA
DEFINITION Sequence 1 from patent US 5710025.
ACCESSION 181227
VERSION 181227.1 GI:3209517
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2556)
AUTHORS Dikstein, R. and Tjhan, R.
TITLE Cell-type specific transcription factor
JOURNAL Patent: US 5710025-A 1 20-JAN-1998;
FEATURES Location/Qualifiers

source 1.2556
BASE COUNT 797 a 614 c 514 g 631 t
ORIGIN

Query Match 99.1%; Score 2534; DB 6; Length 2556;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GGGACCCGTGTGACCAAGTGGTCCGGTCAGCGCCCTCTAAAGTCAGACGCGCCCT 60
DB 1 GGGACCCGTGTGACCAAGTGGTCCGGTCAGCGCCCTCTAAAGTCAGACGCGCCCT 60
QY 61 AGGCTGCTGCTCTCTGAGATGTCGCCGTAAAGCCCCCAACACGACGACATTCAGTTT 120
DB 61 AGGCTGCTGCTCTCTGAGATGTCGCCGTAAAGCCCCCAACACGACGACATTCAGTTT 120
QY 121 CTTGCTAATTTTGAAGCTTCTCCAGGAACGTTTGTATTAAGTAACAGTGGTCCGTTG 180
DB 121 CTTGCTAATTTTGAAGCTTCTCCAGGAACGTTTGTATTAAGTAACAGTGGTCCGTTG 180
QY 181 ATGTGCTATCTCTCTGACCAAGTGTAAAGAGCGAGACCAAGTAACATTAACCTCA 240
DB 181 ATGTGCTATCTCTCTGACCAAGTGTAAAGAGCGAGACCAAGTAACATTAACCTCA 240
QY 241 AGGCGACAGTACAGGAGATCTCTCAACAGTCAAAATCTGTACAGTGGCGAACTTAC 300
DB 241 AGGCGACAGTACAGGAGATCTCTCAACAGTCAAAATCTGTACAGTGGCGAACTTAC 300
QY 301 TCACAAATTAATCAAGAAAGTGGAGAGTACACCTGTTAAAAATTTGGCACAATAGGACT 360
DB 301 TCACAAATTAATCAAGAAAGTGGAGAGTACACCTGTTAAAAATTTGGCACAATAGGACT 360
QY 361 ACTGTGTAACACAGTGTCCGAAGCCTTCTCAGTACATCTGTGGCTGTCCAAACAGT 420
DB 361 ACTGTGTAACACAGTGTCCGAAGCCTTCTCAGTACATCTGTGGCTGTCCAAACAGT 420
QY 421 RGTGCTACAGTACACCTGGAAGAGCCATGTATACGTACATACCTGTGAAGCCTTCAAG 480
DB 421 -GTGCTACAGTACACCTGGAAGAGCCATGTATACGTACATACCTGTGAAGCCTTCAAG 480
QY 481 TTTGGAGAGTACATCCACTCTCTCAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 540
DB 481 TTTGGAGAGTACATCCACTCTCTCAATGAGCCCAATCTTAAAGCAGAGAACTCAGAGC 540
QY 541 TGTTCAGATTAATCTTTCTCCGACAAATGTAGAAATGTGAAGAAATGCAAGAACTTCT 600
DB 541 TGTTCAGATTAATCTTTCTCCGACAAATGTAGAAATGTGAAGAAATGCAAGAACTTCT 600
QY 601 TCGAATGTTAATAAACTAGCATGTAGTATGATACAGTCCCTGGAATGGGGCAAAATGT 660
DB 601 TCGAATGTTAATAAACTAGCATGTAGTATGATACAGTCCCTGGAATGGGGCAAAATGT 660
QY 661 GAAGAAGCTGTGGAACAACTTTTGTGATGCAAAATGGAAGAGAGAAATTTACTAGAA 720
DB 661 GAAGAAGCTGTGGAACAACTTTTGTGATGCAAAATGGAAGAGAGAAATTTACTAGAA 720
QY 721 ACTGTATGTTGAACCAAGTCTTACCTCAGCTCAGCTGCTGTTCTTTCTTAAGAAAG 780
DB 721 ACTGTATGTTGAACCAAGTCTTACCTCAGCTCAGCTGCTGTTCTTTCTTAAGAAAG 780
QY 781 CCGTGTGCTTACGACAACTTCTGCTTACCTCAGAGAGTTCATCAGAAATGTGTCA 840
DB 781 CCGTGTGCTTACGACAACTTCTGCTTACCTCAGAGAGTTCATCAGAAATGTGTCA 840
QY 841 GCAGACTTCTAGTACATGTGTCATGCTTACCTTACATCAACAGTAACACTTCTCTGT 900
DB 841 GCAGACTTCTAGTACATGTGTCATGCTTACCTTACATCAACAGTAACACTTCTCTGT 900
QY 901 GGTGACAACTACAGTGTCTCAAGCCAGTGTGAAGAAATTAATTTCTGTGAGAAC 960
DB 901 GGTGACAACTACAGTGTCTCAAGCCAGTGTGAAGAAATTAATTTCTGTGAGAAC 960
QY 900 GGTGACAACTACAGTGTCTCAAGCCAGTGTGAAGAAATTAATTTCTGTGAGAAC 999

QY 961 AGCACCAGAACTGTGTCACTGTGCAAACTTTGAACCCACTTGTGTGTCAGTGGGACAA 1020
DB 961 AGCACCAGAACTGTGTCACTGTGCAAACTTTGAACCCACTTGTGTGTCAGTGGGACAA 1020
QY 1021 AGCTGAGTGTGTGACACTTCAATCTGTGGGCCCCAAGTGGGCAACAGGGAACAACAG 1080
DB 1021 AGCTGAGTGTGTGACACTTCAATCTGTGGGCCCCAAGTGGGCAACAGGGAACAACAG 1080
QY 1081 TGGAACTGTGTGTGCTTCAAGCTTCAAAACCACTTGTGACATCTGTGGCAACAGTAC 1140
DB 1081 TGGAACTGTGTGTGCTTCAAGCTTCAAAACCACTTGTGACATCTGTGGCAACAGTAC 1140
QY 1141 CACGCTCTACCTCAACCTGAAAGAGCTGTCTGTGAACAGCACTGATCTCTCT 1200
DB 1141 CACGCTCTACCTCAACCTGAAAGAGCTGTCTGTGAACAGCACTGATCTCTCTCT 1200
QY 1201 TCCAGAGTAACTTTTGGAAACCTCAGGTGAGCTATTTGTCTTCCATCTGTGAAAC 1260
DB 1201 TCCAGAGTAACTTTTGGAAACCTCAGGTGAGCTATTTGTCTTCCATCTGTGAAAC 1260
QY 1261 TGTGTTCTCTCTGCTGGGACACATCTGCAAGCTGTATTTGGGAGTCCAGTTCAAT 1320
DB 1261 TGTGTTCTCTCTGCTGGGACACATCTGCAAGCTGTATTTGGGAGTCCAGTTCAAT 1320
QY 1321 CAAGCTTGGCCAGCGGCGCTGTCTTTCACACAGCTGGGATTCACACAGCAGTTTC 1380
DB 1321 CAAGCTTGGCCAGCGGCGCTGTCTTTCACACAGCTGGGATTCACACAGCAGTTTC 1380
QY 1381 AAGCAAGCACTATTTCTCATTTGTTTCAAGTGTGACAGGCTTCAAGGAGCAATGAAA 1440
DB 1381 AAGCAAGCACTATTTCTCATTTGTTTCAAGTGTGACAGGCTTCAAGGAGCAATGAAA 1440
QY 1441 ACAAGTACCACTATTTCTCATTTGTTTCAAGTGTGACAGGCTTCAAGGAGCAATGAAA 1500
DB 1441 ACAAGTACCACTATTTCTCATTTGTTTCAAGTGTGACAGGCTTCAAGGAGCAATGAAA 1500
QY 1501 GATGCGAGTGAACCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
DB 1501 GATGCGAGTGAACCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
QY 1561 AATTACCTGCTGCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
DB 1561 AATTACCTGCTGCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
QY 1621 GAATTAAGAGATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
DB 1621 GAATTAAGAGATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
QY 1681 CTATGCGAGGGGTCAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
DB 1681 CTATGCGAGGGGTCAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
QY 1741 TGTGTCGACACTCATTCAGTCACTGTAAGATGTAAGATGTAAGATGTAAGATGTAAG 1800
DB 1741 TGTGTCGACACTCATTCAGTCACTGTAAGATGTAAGATGTAAGATGTAAGATGTAAG 1800
QY 1801 AAAAGAAATCTTAAGCACTTGTGTAAGAAAGCATGACATTAAGAACTTAATCTGATG 1860
DB 1801 AAAAGAAATCTTAAGCACTTGTGTAAGAAAGCATGACATTAAGAACTTAATCTGATG 1860
QY 1861 TGAATCTGATCTCCCAAGCAACAGAGAGAGTACAGAGGCTTGTGAAGAAAGTGA 1920
DB 1861 TGAATCTGATCTCCCAAGCAACAGAGAGAGTACAGAGGCTTGTGAAGAAAGTGA 1920
QY 1921 CAATGCTGACATCAAGTACATTAAGCAAGCAAGTGAAGTGAAGTGAAGTGAAGTGA 1980
DB 1921 CAATGCTGACATCAAGTACATTAAGCAAGCAAGTGAAGTGAAGTGAAGTGAAGTGA 1980
QY 1981 ATACAGGTTCACAGTCAATTTCTTGAAGAGTGTGATCAATTTGAAGAGCAGAGAAAG 2040
DB 1981 ATACAGGTTCACAGTCAATTTCTTGAAGAGTGTGATCAATTTGAAGAGCAGAGAAAG 2040
QY 2041 AATTGGAAGAAAGAAATGTTACTTAAGGACCAAGAGTGTCTTAAATTAAGAAAGATC 2100

reb 20 16:36:01 2003

us-09-763-909-1.rge

Db 2039 ATTTGGAAGAAAGAAATGTTACTTAAAGCAGCAAGAGCTGTTCTAATTAAGAAAGATC 2098
 QY 2101 CAGAACGCTGAGATTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGAACTTGCACAGA 2160
 Db 2099 CAGAACGCTGAGATTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGAACTTGCACAGA 2158
 QY 2161 TACAGCATAGAGAGCTTAATCTCAGCTCTTGAGCTATTGGACCAAGAAAGAGAC 2220
 Db 2159 TACAGCATAGAGAGCTTAATCTCAGCTCTTGAGCTATTGGACCAAGAAAGAGAC 2218
 QY 2221 CACTAGATCTGCAATTTAGAGGCTTAAAGCAACCTTTGCTTGGGACATCCAGCC 2280
 Db 2219 CACTAGATCTGCAATTTAGAGGCTTAAAGCAACCTTTGCTTGGGACATCCAGCC 2278
 QY 2281 TGACAGCCACCAACAGTTCATCTGCAAGAAATCAGAGAAATCTGCCAGGAGCTTGA 2340
 Db 2279 TGACAGCCACCAACAGTTCATCTGCAAGAAATCAGAGAAATCTGCCAGGAGCTTGA 2338
 QY 2341 TATTTGTATGGAAGAGAGAGAGAGATTTCTCAGCTCTTACCTGGCCCTTC 2400
 Db 2339 TATTTGTATGGAAGAGAGAGAGATTTCTCAGCTCTTACCTGGCCCTTC 2398
 QY 2401 TGAAGTGACCACTCCATCTTCATCCACATCCCTGCTATTTACTGCCAAGAGACACA 2460
 Db 2399 TGAAGTGACCACTCCATCTTCATCCACATCCCTGCTATTTACTGCCAAGAGACACA 2458
 QY 2461 AACCATTTGTCACCTGCTCTGAAATTTCAATTTCTGGAATAACACCAACATGAAAGAG 2520
 Db 2459 AACCATTTGTCACCTGCTCTGAAATTTCAATTTCTGGAATAACACCAACATGAAAGAG 2518
 QY 2521 CATTTGTATGGAAGAGAGAGATTTCTCAGCTCTTACCTGGCCCTTC 2558
 Db 2519 CATTTGTATGGAAGAGAGAGATTTCTCAGCTCTTACCTGGCCCTTC 2556

RESULT 2
 HSTAFII
 LOCUS HSTAFII 2556 bp mRNA linear PRI 14-NOV-1996
 DEFINITION H.sapiens TAFII105 mRNA, partial.
 ACCESSION Y09321
 VERSION Y09321.1 GI:1669688
 KEYWORDS TAFII105 gene.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 2556)
 Dikstein,R., Zhou,S. and Tjian,R.
 Human TAFII 105 is a cell type-specific TFIID subunit related to
 hTAFII30
 JOURNAL Cell 87 (1), 137-146 (1996)
 MEDLINE 97011146
 PUBMED 8858156
 REFERENCE 2 (bases 1 to 2556)
 AUTHORS Dikstein,R.
 TITLE Direct Submission
 JOURNAL Submitted (07-NOV-1996) R. Dikstein, Weizmann Institute of Science,
 Dept. Biochemistry, Rehovot, 76100, ISRAEL
 FEATURES
 source
 1..2556
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Hela"
 1..2406
 /gene="TAFII105"
 <1..2406
 /gene="TAFII105"
 /codon_start=1
 /product="TBP associated factor"
 /protein_id="CAA70499.1"
 /db_xref="GI:1669689"
 /db_xref="SWISS-PROT:O92750"

/translation="GTLVTKVAPVSPKVSQRLPAQIVAKAPNTTIOEPANL
 QLEPGTVLILKSNGLPLVSPQFTVRAETSNITSRPAPRPOVRICTVPSNSO
 LKKAVATPVKRLAIGTVVTPKPSQSVQAVPTSVTVTPGRLNTVTLKRS
 LKASPSNEPINKAKNSAANOINLSPMLENKKKKNLMLILKAGSGSSPEQO
 NKRYLEQLIDAKTEAEETRLKLYELKSSPOHVLKFKLSVALROLIPNSQFIO
 QCVQPTSSDMVLAICTVTYVTSVTVVSSQSEKSIIVSGATAPRVSQVNLPLA
 GPVAKAGVYVTLHSGVPTAATGGTATGTLQTSKPLVSNANTVTVSLOPEKPVLS
 GYAVVTLSPAVTEGSGAALCLPSVKPVSFCMDHICKPVIGTPQIKLAQGPVLS
 OPAGIPTSSKQLFSLFHYVQOPSGSNKQVYTSHSSTLTIOKCGQTMPTVITP
 TQOPFPASTIKOTLPLPKNKLISLQASPTORNKIKENYSCFREDIDINVTSAVNL
 NEBNACTLITNSLVTLLQSCDEPFLTGALQKRLIDIGKHDTIELNSAVNLIS
 QATQERLRLKLEKTLAIOHRMTYKASBNYILICSDPSQLFLEKLDLEOKRLE
 ERMELKKAASSNSKEDPEQLRLKOKAKLEQLOIHRPANTLALALIGPKRKP
 LKSGIEGLKNDNLASGTSLSLFTAKQLHRPRITRICLDLIFCMEQREKYSBALYLA
 LK"

BASE COUNT 797 a 614 c 514 g 631 t
 ORIGIN

Query Match 99.1%; Score 2534; DB 9; Length 2556;
 Best local Similarity 99.9%; Pred. No. 0;
 Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
 QY 1 GGGACCTGGTGACCAAGTGGCTCGGTCAGCGCCCTCTTAAGTCACAGCGGCTT 60
 Db 1 GGGACCTGGTGACCAAGTGGCTCGGTCAGCGCCCTCTTAAGTCACAGCGGCTT 60
 QY 61 AGGCTGCTGCTCCTCAGATAGTGGCGGTGAAGCCGCCAAGCCAGCAATCCATTT 120
 Db 61 AGGCTGCTGCTCCTCAGATAGTGGCGGTGAAGCCGCCAAGCCAGCAATCCATTT 120
 QY 121 CTTGCTAATTTGGAGCTTCCCTCCAGAAACCGTTTGTATTAAGTAACAGTGGCTGG 180
 Db 121 CTTGCTAATTTGGAGCTTCCCTCCAGAAACCGTTTGTATTAAGTAACAGTGGCTGG 180
 QY 181 ATGTGCTATCTCTTCAAGCAACTGTAAACAGAGCCGAGACCAAGTAACATTAACCTCA 240
 Db 181 ATGTGCTATCTCTTCAAGCAACTGTAAACAGAGCCGAGACCAAGTAACATTAACCTCA 240
 QY 241 AGGCCAGCAGTACACAGGAACTCCTCAACAGCTGAATAATGTGTAATGCCGAACCTAGC 300
 Db 241 AGGCCAGCAGTACACAGGAACTCCTCAACAGCTGAATAATGTGTAATGCCGAACCTAGC 300
 QY 301 TCACATTAATCAAGAAAGTGCAGTACACCTGTTAAATAATGGCACAAATAGAACT 360
 Db 301 TCACATTAATCAAGAAAGTGCAGTACACCTGTTAAATAATGGCACAAATAGAACT 360
 QY 361 ACTGTGTAACCACTGTTCCGAAGCTTCTCAGTAAGTAATGTGCTGTGCCAACAGT 420
 Db 361 ACTGTGTAACCACTGTTCCGAAGCTTCTCAGTAAGTAATGTGCTGTGCCAACAGT 420
 QY 421 RGTGTCACAGTACTCTGGAAGGCAATGATGTAATGTAATGTAATGTAATGTAATGTAATG 480
 Db 421 -GTGTCACAGTACTCTGGAAGGCAATGATGTAATGTAATGTAATGTAATGTAATGTAATG 479
 QY 481 TTTGGAGCATCATCCACTCTTCAATGAGCCCAATTTAAAGCAGAGAACTCAGCAGC 540
 Db 481 TTTGGAGCATCATCCACTCTTCAATGAGCCCAATTTAAAGCAGAGAACTCAGCAGC 539
 QY 541 TGTTCAGATTAATCTTCTCCGCAATGCTAGAAATGTGAAGAAATGCAAGAACTTCT 600
 Db 541 TGTTCAGATTAATCTTCTCCGCAATGCTAGAAATGTGAAGAAATGCAAGAACTTCT 599
 QY 601 TGCATGTTAATAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 Db 601 TGCATGTTAATAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 659
 QY 661 GAAGAAGCTGTGGAACAATTTGGATGCAAAAAAGCAGAGAGAAATTAAGTAGAA 720
 Db 661 GAAGAAGCTGTGGAACAATTTGGATGCAAAAAAGCAGAGAGAAATTAAGTAGAA 719
 QY 721 ACTGATGTTGAAGTCAAGTCTTCACTCAGCTGCTGTTCTTTCTTAAGAAAG 780
 Db 721 ACTGATGTTGAAGTCAAGTCTTCACTCAGCTGCTGTTCTTTCTTAAGAAAG 779

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 21.8929 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GTLVTKAPVAPSPKVS...KZHOHERALFTITLTLTY 852

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3990	93.6	801	1 US-08-725-012-2	Sequence 2, Appli
2	1307.5	30.7	737	1 US-08-188-582-16	Sequence 16, Appl
3	1307.5	30.7	737	1 US-08-646-715-16	Sequence 16, Appl
4	775	18.2	921	1 US-08-188-582-2	Sequence 2, Appli
5	775	18.2	921	1 US-08-646-715-2	Sequence 2, Appli
6	237	5.6	2035	1 US-08-046-585-5	Sequence 5, Appli
7	237	5.6	2035	1 US-08-393-703-5	Sequence 5, Appli
8	237	5.6	2035	5 PCT-US93-11721-5	Sequence 5, Appli
9	198	4.6	862	1 US-08-325-267A-4	Sequence 5, Appli
10	197	4.6	2972	1 US-08-579-181-2	Sequence 4, Appli
11	197	4.6	3118	4 US-09-579-181-1	Sequence 2, Appli
12	191.5	4.5	894	3 US-08-362-525-22	Sequence 1, Appli
13	191.5	4.5	894	3 US-08-971-692-15	Sequence 22, Appl
14	184.5	4.3	1537	1 US-08-325-267A-2	Sequence 15, Appl
15	178	4.2	1721	3 US-08-700-651-5	Sequence 2, Appli
16	178	4.2	1721	3 US-08-928-361B-6	Sequence 5, Appli
17	175.5	4.1	3969	4 US-08-061-316-5	Sequence 6, Appli
18	172	4.0	750	4 US-09-165-239A-4	Sequence 5, Appli
19	168	3.9	1837	3 US-08-928-361B-5	Sequence 5, Appli
20	166	3.9	805	4 US-09-103-429A-4	Sequence 4, Appli
21	164.5	3.9	752	1 US-08-244-189-2	Sequence 4, Appli
22	164	3.8	903	2 US-08-853-310-2	Sequence 2, Appli
23	160	3.8	1125	4 US-09-513-783A-152	Sequence 2, Appli
24	160	3.8	1610	4 US-09-513-783A-22	Sequence 12, Appl
25	158.5	3.7	786	4 US-09-103-429A-3	Sequence 3, Appli
26	157.5	3.7	941	4 US-07-757-022B-14	Sequence 14, Appl
27	157.5	3.7	1022	4 US-07-757-022B-84	Sequence 84, Appl

28	157.5	3.7	1038	4 US-07-757-022B-74	Sequence 74, Appl
29	157.5	3.7	1049	4 US-07-757-022B-58	Sequence 58, Appl
30	157.5	3.7	1140	4 US-07-757-022B-104	Sequence 104, Appl
31	157.5	3.7	1270	4 US-07-757-022B-44	Sequence 44, Appl
32	157.5	3.7	1311	4 US-07-757-022B-42	Sequence 42, Appl
33	157.5	3.7	1313	4 US-07-757-022B-142	Sequence 142, Appl
34	157.5	3.7	1314	4 US-07-757-022B-50	Sequence 50, Appl
35	157.5	3.7	1320	4 US-07-757-022B-46	Sequence 46, Appl
36	157.5	3.7	1354	4 US-07-757-022B-60	Sequence 60, Appl
37	157.5	3.7	1361	4 US-07-757-022B-48	Sequence 48, Appl
38	157.5	3.7	1361	4 US-07-757-022B-40	Sequence 40, Appl
39	157.5	3.7	1363	4 US-07-757-022B-52	Sequence 52, Appl
40	157.5	3.7	1404	4 US-07-757-022B-2	Sequence 2, Appli
41	157.5	3.7	1404	4 US-07-757-022B-62	Sequence 62, Appl
42	157	3.7	2843	1 US-07-741-940-2	Sequence 2, Appli
43	157	3.7	2843	1 US-08-289-548A-2	Sequence 2, Appli
44	157	3.7	2843	1 US-08-452-654-2	Sequence 2, Appli
45	157	3.7	2843	2 US-08-370-235A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-725-012-2
Sequence 2, Application US/08725012
Patent No. 5710025

GENERAL INFORMATION:

APPLICANT: Dikstein, Rivka
APPLICANT: Tjian, Robert
TITLE OF INVENTION: B-Cell Specific Transcription Factor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012

FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-012-2

Query: Match 93.6% Score 3990: DB 1: Length 801:
Best Local Similarity 99.9% Pred. No. 9.6e-319;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTLVTKAPVAPSPKVS...KZHOHERALFTITLTLTY 852
DB 1 GTLVTKAPVAPSPKVS...KZHOHERALFTITLTLTY 852
OY 61 MLVSPQQTTRAEFTSNITSPRAVPANFQYVICTVPSSQLIKKAVPEVKRLAIGT 120

```

Db      61  MLYSPQOTVTRAEITSNITSRPVPANPQVTKICTVBNSSQLIKRVAVTPVKLAQIGT 120
QY      121  TVVTVPKPSSVQSAVPTSVTVTPGKPNVTYTTTLKPSLSGASSTPSNPNKAESAA 180
Db      121  TVVTVPKPSSVQSAVPTSVTVTPGKPNVTYTTTLKPSLSGASSTPSNPNKAESAA 180
QY      181  VOINLSPTMLENKKCNFLAMLIKACSSQSPSEMGOVKKLVEQLLDKAEFEPTK 240
Db      181  VOINLSPTMLENKKCNFLAMLIKACSSQSPSEMGOVKKLVEQLLDKAEFEPTK 240
QY      241  LVELKSSPQPHLVPLFKSVVALROLPLPSOSFIOQCVOQTSDDMYATCTTVTTSVP 300
Db      241  LVELKSSPQPHLVPLFKSVVALROLPLPSOSFIOQCVOQTSDDMYATCTTVTTSVP 300
QY      301  VTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTHSVGPATAGGTGA 360
Db      301  VTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTHSVGPATAGGTGA 360
QY      361  GGLGLOTSRPLVTSVANTYTTVTSLOPEKPVSGTAVTSLPVTGEGTGAICLPSVP 420
Db      361  GGLGLOTSRPLVTSVANTYTTVTSLOPEKPVSGTAVTSLPVTGEGTGAICLPSVP 420
QY      421  VVSFCMDHICKPVIGPVOIKLAOPBPVLSOPAGIPTGSSSKOLPSLFHVVOQPSGNEK 480
Db      421  VVSFCMDHICKPVIGPVOIKLAOPBPVLSOPAGIPTGSSSKOLPSLFHVVOQPSGNEK 480
QY      481  QVTTSHSSTLTLOKCGOKTNPVNTIIPTSOPPPASITLKOITLPKNTLISLASPTOKR 540
Db      481  QVTTSHSSTLTLOKCGOKTNPVNTIIPTSOPPPASITLKOITLPKNTLISLASPTOKR 540
QY      541  IKENTSCRRDDINDVTSMAGVNLNENACILATNSELVGTLLIOSCDEPFILGALO 600
Db      541  IKENTSCRRDDINDVTSMAGVNLNENACILATNSELVGTLLIOSCDEPFILGALO 600
QY      601  KRILDIGKKHDTLNSDAVNLISQATQERLGLLEKTLAIOHMTYTKASENYILCSO 660
Db      601  KRILDIGKKHDTLNSDAVNLISQATQERLGLLEKTLAIOHMTYTKASENYILCSO 660
QY      661  TRSOLKFEKLDQLEKORRDEEREMELKAASRSNKEDPEQLRIKOKAKELQOELAOI 720
Db      661  TRSOLKFEKLDQLEKORRDEEREMELKAASRSNKEDPEQLRIKOKAKELQOELAOI 720
QY      721  QHRANLATAAIGRRKRPLESGIEGLKDNILASGSSSLATKQOLHRRPRTIRICLDLI 780
Db      721  QHRANLATAAIGRRKRPLESGIEGLKDNILASGSSSLATKQOLHRRPRTIRICLDLI 780
QY      781  FCMOEREMKYSRALYIALLK 801
Db      781  FCMOEREMKYSRALYIALLK 801

```

RESULT 2

US-08-188-582-16
Sequence 16, Application US/08188582
Patent No. 5534410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
NUMBER OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA

```

ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-16

```

Query Match 30.7%; Score 1307.5; DB 1; Length 737;
Best Local Similarity 40.0%; Pred. No. 1,3e-98;
Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;

```

QY      13  PKVSSG-----PRLPAQIVAKAPNTTIOEPANLOLPGTVILKSSGLM 61
Db      68  PLATSGIRATLPVPLRLPQ-----PONTNIO---NFQLPQAVLRSNGQL 118
QY      62  LVSPOQVTR-----ATTSNITSRPVPANPQVTKICTVBNSSQLIKRVAVTPVKL 115
Db      119  MI-PQOLAQOAOAHAPQPTMAPRATPTSAVPQISTVQAPQPTIAR-QVTP----- 172
QY      116  AQIGTVTVTPKPSVQSAVPTSVTVTPGKPLNVT--TLKPSLSGASS-----TPS 168
Db      173  ---TTIKOV---SQAOITVQPSATLQKSPGVOPOLVGGAAQTASIGTAVQGTGPQ 225
QY      169  NE-PRLAENSAAVOINLSPTMLENKKCNFLAMLIKACSSQSPSEMGOVKKIYBOL 227
Db      226  RTVPGATTTSSAATE-----TMENVKCKNFLTILIKIASSGQSTETANVELVONT 279
QY      228  LDKIEAEFEPTKLYVELKSSPQPHLVPLFKSVVALROLPLPSOSFIOQCVOO-----TS 283
Db      280  LDGKIEADEFTSRILYRELNSSPOPYLVPLKRSIPALROLPPDSARIQSQQPPPTTS 339
QY      284  SDMYATCTTVTTSVPVTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
Db      340  Q-----ATTALTAVALVSSVQRTACKTAAYTSALQPVLSL----- 376
QY      344  VTLHSGVPLAATGGTAGTGLQTSKPLVTSVANTYTTVTSLOPEKPVSGTAVTSLP 403
Db      377  -----TPQGVGKQCGQPPPLVIG-----QPKP-----GALIRPPQV 410
QY      404  TFGTSGAICLPSVPRVVSFCMDHICKPVIGPVOIKLAOPBPVLSOPAGIPTGSSSKO 463
Db      411  TLQOT-----PMVALRQPH-NRIMLTTPQOIQL----- 437
QY      464  LFLSLFHVVOQPSGNEKQVTTISHSSTLTLOKCGOKTNPVNTIIPTSOPPPASITL 533
Db      438  ----- 511
QY      524  PGKNTI---SLQSPQOKRIKENYTCFRDEDDINDVTSMAGVNLNENACILATNSEL 580
Db      452  PGTKALSAVSAQAAQAKKIKKEGGGSRDDDDINDVASMAGVNLSESRATLATNSEL 511
QY      581  VGLTLOSCDEPFILGALQOKRLDIGKRDITELNSDAVNLISQATQERLGLLEKTLA 640
Db      512  VGLTFRSCDEFTLLQAPLQORILIEIGKKGITELHPDVVSVYSHATQORLQINVEKISE 571

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 16.4197 Seconds
(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GTLVTKVAPVAPKVSQSP.....KZHOHERALFTIRLLTLTV 852

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3990	93.6	801	T2DT_HUMAN	Q92750 homo sapien
2	1307.5	30.7	1083	T2D3_HUMAN	000268 homo sapien
3	775	18.2	921	T2D3_DROME	P47825 drosophila
4	239.5	5.6	2035	HFCL_HUMAN	P51610 homo sapien
5	238	5.6	2090	HFCL_MESAU	P51611 mesocricetu
6	220	5.2	1367	AMTH_YEAST	P08640 saccharomyc
7	219	5.1	5179	MUC2_HUMAN	002817 homo sapien
8	213.5	5.0	5376	ZAN_MOUSE	088799 mus musculu
9	206.5	4.8	865	CPN_DROME	002910 drosophila
10	205.5	4.8	1161	DAN4_YEAST	P47179 saccharomyc
11	203.5	4.8	1322	YAG3_YEAST	P39712 saccharomyc
12	201	4.7	725	AGAL_YEAST	P32323 saccharomyc
13	199	4.7	2700	ZAN_HUMAN	Q9Y493 homo sapien
14	197.5	4.6	881	YTH8_YEAST	P47033 saccharomyc
15	197.5	4.6	1609	FIG2_YEAST	P25653 saccharomyc
16	193.5	4.5	670	VG50_HSV1	000130 ictaluriid b
17	191.5	4.5	1260	ALSL1_CANAL	P46590 candida alb
18	191	4.5	797	VGIX_HAYEB	P28968 equine hepr
19	187.5	4.4	3726	TRX_DROME	P20659 drosophila
20	186	4.4	3178	YS89_CAEEL	Q09624 caenorhabdi
21	184.5	4.3	1537	FLOI_YEAST	P32768 saccharomyc
22	182	4.3	1075	FLOS_YEAST	P38894 saccharomyc
23	182	4.3	3866	HRX_MOUSE	P55200 mus musculu
24	181.5	4.3	662	MUC1_XENLA	Q05049 xenopus lae
25	181.5	4.3	1199	N121_RAT	P52591 rattus norv
26	181	4.2	1858	P3K2_DICDI	P54674 dictyosteli
27	177.5	4.2	1140	YH96_YEAST	Q04893 saccharomyc
28	176	4.1	2541	TAL1_HUMAN	Q9Y490 homo sapien
29	175.5	4.1	2090	N214_HUMAN	P33658 homo sapien
30	175.5	4.1	3969	HRX_HUMAN	Q03164 homo sapien
31	173.5	4.1	1802	HKR1_YEAST	P41809 saccharomyc
32	173	4.1	2404	SON_MOUSE	Q9GX47 mus musculu
33	172	4.0	886	VGP3_EBVA8	Q07284 epstein-bar

34	172	4.0	1018	1	HMW1_MYCPN	Q50365 mycoplasma
35	172	4.0	1119	1	AL53_CANAL	Q74623 candida alb
36	172	4.0	1169	1	YK82_YEAST	P36170 saccharomyc
37	171.5	4.0	1780	1	YK26_CAEEL	P34333 caenorhabdi
38	170	4.0	1306	1	MSB2_YEAST	P32334 saccharomyc
39	169.5	4.0	388	1	MPRT_YEAST	P50105 saccharomyc
40	169.5	4.0	745	1	OCT1_PIG	Q29076 sus scrofa
41	169.5	4.0	2426	1	SON_HUMAN	P18583 homo sapien
42	169.5	4.0	1229	1	N121_HUMAN	Q9Y293 homo sapien
43	166.5	3.9	743	1	OCT1_HUMAN	P14859 homo sapien
44	165	3.9	606	1	SP2_HUMAN	Q02086 homo sapien
45	165	3.9	2541	1	TAL1_MOUSE	P26039 mus musculu

ALIGNMENTS

```

RESULT 1
ID T2DT_HUMAN STANDARD: PRT: 801 AA.
AC Q92750:
DF 15-JUL-1998 (Rel. 36, Created)
DF 15-JUL-1998 (Rel. 36, Last sequence update)
DF 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIID 105 kDa subunit (TAFI1-105)
DE (TAFII105) (Fragment).
GN TAFAB OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=97011146; PubMed=8858156;
RX Diktstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIID subunit related to
RT hTAFII130."
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
CC -1- A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
CC MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
CC PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
CC HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y09321; CA70499.1;
CC Genew: HGNC:11538; TAFAB.
CC MIM: 601689;
CC DR InterPro: IPR003894; TAF_hom.
CC DR SMART: SM00549; TAFH; 1.
CC KW Transcription regulation; Nuclear protein.
CC FT NON_TER
CC SQ SEQUENCE 801 AA; 85558 MW; D12B4932EFA9CD2 CRC64;

Query Match 93.6%; Score 3990; DB 1; Length 801;
Best Local Similarity 99.9%; Pred. No. 5.1e-199;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 GTLVTKVAPVAPKVSQSPRLAPQIVAVKAPNTTIOFPANLQLPCTVLIKNSGPI 60
|||||

```

Db 1 GTLVTKAVPAPVAPKSSGRLPAPQIVAVKAPNTTIIQFPANQLPFGTVLISNSGPL 60
 QY 61 MLYSPQOTVTRAEFTSNITSRPAPVANDQVYKICTVPSNSQLIKKVAVTPVKRLAQIGT 120
 Db 61 MLYSPQOTVTRAEFTSNITSRPAPVANDQVYKICTVPSNSQLIKKVAVTPVKRLAQIGT 120
 QY 121 TVYTVTRKPSVSVAVPTSVYTVTPKRPPLNTVYTLKRPSSIGASTSNENKLAENSA 180
 Db 121 TVYTVTRKPSVSVAVPTSVYTVTPKRPPLNTVYTLKRPSSIGASTSNENKLAENSA 180
 QY 181 VOJNLSPMLNENKCKKNFLAMLKIKLACSGSGSPGEMGONVYKYLEOLDAKIEAEFTPK 240
 Db 181 VOJNLSPMLNENKCKKNFLAMLKIKLACSGSGSPGEMGONVYKYLEOLDAKIEAEFTPK 240
 QY 241 LVEELKSPQPHLPVLPKRSVVALROLPLNSQSFIOOCVOYOTSSDMYIATCTTVTTSV 300
 Db 241 LVEELKSPQPHLPVLPKRSVVALROLPLNSQSFIOOCVOYOTSSDMYIATCTTVTTSV 300
 QY 301 VTTVSSSSQSEKSTIVSGARPRVSVQTLNPLAGPVGAKAGVYTLHSVGTATGCTTA 360
 Db 301 VTTVSSSSQSEKSTIVSGARPRVSVQTLNPLAGPVGAKAGVYTLHSVGTATGCTTA 360
 QY 361 GTGLQTSKPLVTSVANTVTVTSIQPEKPVYSGTAATLSLPAYTFGETSGAATLPSVKP 420
 Db 361 GTGLQTSKPLVTSVANTVTVTSIQPEKPVYSGTAATLSLPAYTFGETSGAATLPSVKP 420
 QY 421 VVSFCMDHCKPVIGTGVQIKLAOPPVLSOPACIPGSSSKQLSLFHVVOQPSGNEK 480
 Db 421 VVSFCMDHCKPVIGTGVQIKLAOPPVLSOPACIPGSSSKQLSLFHVVOQPSGNEK 480
 QY 481 QVTIHSSTLTIOKCGOKMPVNTIIPTSQFPASITKOTLPGNLTLSQASPTOKNR 540
 Db 481 QVTIHSSTLTIOKCGOKMPVNTIIPTSQFPASITKOTLPGNLTLSQASPTOKNR 540
 QY 541 IKENWTSQFREDINDVTSAGVNLNENACILATNSELVGLTIOGCKDEPFETALQ 600
 Db 541 IKENWTSQFREDINDVTSAGVNLNENACILATNSELVGLTIOGCKDEPFETALQ 600
 QY 601 KRILIDKCKHDITELNSDAVNLISOAQOERLGLLEKTLTAIOHRMTYKASENYILCS 660
 Db 601 KRILIDKCKHDITELNSDAVNLISOAQOERLGLLEKTLTAIOHRMTYKASENYILCS 660
 QY 661 TRSOLKLEKIDOLEKOKDEEREMLKAKSRNKEDDEPOLKOKAKELQOLEAQI 720
 Db 661 TRSOLKLEKIDOLEKOKDEEREMLKAKSRNKEDDEPOLKOKAKELQOLEAQI 720
 QY 721 QHRDANLTATAAIGPRKRPESGIEGLKDNLLASGTSSTLATKQLRPRITRICLDLI 780
 Db 721 QHRDANLTATAAIGPRKRPESGIEGLKDNLLASGTSSTLATKQLRPRITRICLDLI 780
 QY 781 FCMQEREMKYSRALTALAK 801
 Db 781 FCMQEREMKYSRALTALAK 801

RESULT 2

T2D3_HUMAN STANDARD; PRT; 1083 AA.
 AC 000268; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription initiation factor TFIIID 135 kDa subunit (TAFII-135)
 GN TAFII135 (TAFII-130) (TAFII130).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=97336072; PubMed=9192867;
 RX Mengus G., May M., Carre L., Chambon P., Davidson I.;

RT "human TAF(II)135 potentiates transcriptional activation by the AF-2s
 of the retinoic acid, vitamin D3, and thyroid hormone receptors in
 mammalian cells.";
 RT Genes Dev. 11:1381-1395(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida A.K., Baggeley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levasseur M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMuray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.L.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay B.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swan R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Yaldin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RT Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=97098442; PubMed=8942962;
 RA Tanese N., Saluja D., Vaasallo M.F., Chen J.-L., Admon A.;
 RT "Molecular cloning and analysis of two subunits of the human TFIIID
 complex: hTAFII130 and hTAFII100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTION
 CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: Y11354; CAA72189.1; -
 CC EMBL: AL137077; CAC36006.1; -
 CC EMBL: AL109911; CAC22312.2; -
 CC EMBL: U75308; AAC50901.1; -
 CC TRANSFAC: T02328; -
 CC GeneW: HGNC:11537; TAF4.
 CC MIM: 601796; -
 CC InterPro: IPR003894; TAF_hom.
 CC SMART: SM00549; TAFH.1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 39 42
 FT DOMAIN 52 57
 FT DOMAIN 98 101
 FT DOMAIN 142 148
 FT DOMAIN 268 275
 FT POLY-HIS.
 FT POLY-ALA.
 FT POLY-GLY.
 FT POLY-ALA.
 FT POLY-PRO.

FT
 FT
 DOM.
 20 16:36

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.